



Le système documentaire Vercingétorix I

Gabriel G. Bès, Pierre-Maurice Fauchère

► To cite this version:

Gabriel G. Bès, Pierre-Maurice Fauchère. Le système documentaire Vercingétorix I. Condenser - Adosa, Clermont-Ferrand, 1980, 1, pp.57-94. hal-01071423v2

HAL Id: hal-01071423

<https://hal.science/hal-01071423v2>

Submitted on 18 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le système documentaire Vercingétorix I

Gabriel G. Bès et Pierre-Maurice Fauchère
Groupe de recherches sur la condensation de l'information en langue naturelle (CILN)
Université Blaise-Pascal, Clermont II

Condenser, Adosa, Clermont-Ferrand, février 1980, n° 1, p. 57-94

Résumé

Vercingétorix I (= V1) est un système documentaire composé (a) d'un langage documentaire utilisé par le documentaliste pour produire les expressions destinées à indexer les documents et à exprimer les questions documentaires, (b) d'une aide au documentaliste et (c) d'un ensemble de procédures informatisées, destinées à stocker les expressions documentaires, analyser les questions documentaires et rechercher les documents pertinents.

V1 se propose de réduire le bruit documentaire, par l'introduction de relations syntaxiques dans les expressions documentaires, et le silence documentaire, en situant chaque expression documentaire dans un système d'interrelations et de généralisations avec les autres expressions documentaires dans le réseau sémantique ; les expressions documentaires doivent donc pouvoir être « comprises » par la machine, qui doit interpréter la relation entre facteurs sémantiques et syntaxiques de chaque expression. Ces expressions documentaires doivent, en outre, pouvoir être lues directement par un être humain, afin de permettre la constitution automatique d'outils bibliographiques tels que bulletins de résumés, index, etc.

V1 est construit en fonction de caractéristiques des langues naturelles et en tenant compte de la fonction de relais du documentaliste.

Le présent texte comprend quatre parties. Le § 1 définit le système et en expose les objectifs et les fondements. Le § 2 distingue les différents types d'information associés au contenu de chaque document : leur mode de production (direct ou indirect, via commutation, permutation ou développement de l'anaphorique), leurs fonctions et leur présentation. Le §3 donne les principes généraux du système d'assistance de l'ordinateur à l'indexation : assistance matérielle, contrôle de la correction des structures et optimisation de l'indexation. Le § 4 donne une dizaine d'exemples d'indexation.

Voir aussi

Gabriel G. Bès et Pierre-Maurice Fauchère. « Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales (1^{re} partie) ». *Condenser*, Adosa, Clermont-Ferrand, janvier 1981, n° 2, p. 39-89. <http://hal.archives-ouvertes.fr/hal-01117883>

Gabriel G. Bès et Pierre-Maurice Fauchère. « Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales (2^e partie) ». *Condenser*, Adosa, Clermont-Ferrand, avril 1982, n° 3, p. 3-31. <http://hal.archives-ouvertes.fr/hal-01117888>

Gabriel G. Bès. « Les structures syntaxiques du langage Vercingétorix I. Problèmes liés à leur naturalité. » *Traitement automatique des langues naturelles et systèmes documentaires*, *Condenser*, Supplément n° 1, Adosa, Clermont-Ferrand, 1984, p. 109-134.
<http://hal.archives-ouvertes.fr/hal-01071426>

NB. La version 2 de ce document déposée dans HAL ne diffère de la version 1 que par les liens ci-dessus, qui ont dû être corrigés suite à une défaillance du système ayant conduit au remplacement de certains dépôts par d'autres n'ayant rien à voir.

LE SYSTEME DOCUMENTAIRE VERCINGETORIX I

La description du système Vercingétorix I prendra la forme de fiches numérotées selon l'ordre de parution et consacrées à chacun des aspects du système. Ces fiches pourront être ultérieurement reprises selon un plan de classement méthodique.

Les considérations plus théoriques et l'examen de la bibliographie afférente figureront sous une rubrique "Linguistique et langages documentaires".



Fiche n°1 (août 1979)

G.G.Bès et P.-M.Fauchère

Vercingétorix 1. Définition, objectifs et fondements

Vercingétorix 1 (= V1) est un système documentaire composé de:

(A) un langage documentaire, constitué d'un lexique, où les termes admis par ce langage et leurs propriétés sont spécifiés, et d'un ensemble de règles syntaxiques. Ce langage est utilisé par le documentaliste pour produire les expressions documentaires (= ED), destinées à indexer les documents et à exprimer les questions documentaires.

(B) une aide au documentaliste, comprenant:

(1) un manuel d'acquisition et d'utilisation, destiné au documentaliste qui doit acquérir et utiliser ce langage;

(2) une assistance par l'ordinateur, dans le cadre d'un système interactif, aussi bien pour l'indexation que pour la recherche des documents pertinents.

(C) un ensemble de procédures informatisées, destinées à:

(1) stocker les ED;

(2) analyser les questions documentaires et les interpréter à différents niveaux de généralisation en constituant un réseau sémantique entre les ED;

(3) rechercher les documents pertinents.

V1 est la version la plus simple d'une série de langages de complexité syntaxique croissante. Il se propose de réduire le bruit documentaire, par l'introduction de relations syntaxiques dans les ED, et le silence documentaire, en situant chaque ED dans un système d'interrelations et de généralisations avec les autres ED dans le réseau sémantique; les ED doivent donc pouvoir être "comprises" par la machine, qui doit interpréter la relation entre facteurs sémantiques et syntaxiques de chaque ED. Ces ED doivent, en outre, pouvoir être lues directement par un être humain, afin de permettre

la constitution automatique d'outils bibliographiques tels que bulletins de résumés, index, etc. Malgré cet ensemble d'exigences, la simplicité de l'acquisition et de l'utilisation par le documentaliste du langage documentaire défini dans V1 doit être soigneusement préservée.

Les objectifs de V1 se situent à deux niveaux: particulier et général. Au niveau particulier, V1 se propose de traiter comme domaine référentiel les sciences de l'information et de la documentation. A un niveau général, on se propose, par le biais de V1, de dégager des structures syntaxiques et sémantiques suffisamment abstraites et générales pour être transposées dans d'autres domaines référentiels.

V1 n'est ni un système d'indexation automatique, ni un langage artificiel comme le Syntol ou le Semantic Code de la Western Reserve University, où les informations des énoncés en langue naturelle doivent être traduites en formules de ce langage.

L'analyse conceptuelle qui sous-tend la préparation de V1 peut se résumer dans les points suivants:

(a) Une analyse automatique du contenu des textes qui permette une indexation automatique satisfaisante est aujourd'hui impossible.

(b) Par conséquent, le documentaliste reste le relais indispensable entre le document original, en langue naturelle, et le langage susceptible d'être traité par la machine.

(c) Le documentaliste "connaît" intuitivement la langue naturelle qu'il utilise.

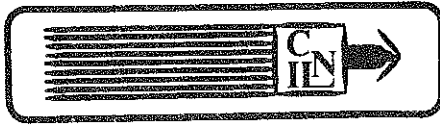
(d) Les langues naturelles possèdent des caractéristiques qu'on commence à connaître suffisamment bien pour pouvoir les utiliser dans la confection des langages documentaires. Ces caractéristiques sont assez générales pour se retrouver sinon dans toutes les langues, du moins dans beaucoup d'entre elles, notamment dans celles qui véhiculent l'essentiel de l'information scientifique et technique.

(e) Un langage documentaire construit en fonction de certaines structures des langues naturelles pourra être appris plus rapidement et manié plus commodément qu'un langage entièrement artificiel.

V1 est donc construit en fonction de caractéristiques des langues naturelles et en tenant compte de la fonction de relais du documentaliste. Toute structure du langage documentaire défini dans V1 est susceptible d'être asso-

ciée à une structure de la langue naturelle utilisée par le documentaliste. V1 doit être acquis et utilisé "naturellement" par celui-ci, c'est-à-dire en faisant référence à son propre langage, qu'il doit "retrouver" dans les ED. Le Manuel d'acquisition et d'utilisation est ainsi une pièce essentielle de V1 (le manuel n'indiquera pas comment comprendre un document original; il suppose que cette compréhension est acquise. Il doit indiquer, en revanche, comment "traduire" l'information contractée du texte original en ED).

La présence généralisée, dans les langues qui véhiculent l'information scientifique et technique, des structures linguistiques utilisées par le langage documentaire défini dans V1, doit permettre à celui-ci, par traduction automatique des ED, d'évoluer vers un système multilingue.



Fiche n°2 (août 1979)

G.G.Bès et P.-M.Fauchère

Les informations indexées

Sommaire:

- 1 Introduction
- 2 Les informations indexées et leur classement
 - 2.1 Procédés d'intervention de l'indexeur
 - 2.1.1 Les instructions d'écriture
 - 2.1.1.1 L'instruction de commutation
 - 2.1.1.2 L'instruction de permutation
 - 2.1.1.3 L'instruction de développement de l'anaphorique
 - 2.2 Les fonctions des informations indexées
 - 2.3 La présentation des informations indexées
 - 3 Les structures lisibles
 - 4 Récapitulation

1 INTRODUCTION

Le système documentaire VI est composé, comme il a été indiqué dans la fiche n°1, d'un langage documentaire, d'une aide au documentaliste, dont une partie est constituée d'une assistance par l'ordinateur, et d'un ensemble de procédures informatisées. Trois entités, le documentaliste, le langage documentaire et la machine, sont donc en relation pour traiter les informations relevant du contenu de chaque document. Par ailleurs, il a été indiqué, sur la même fiche n°1, que les expressions documentaires (au sens large du terme; cf. ci-dessous, au § 2.3, des précisions sur cette notion) doivent être "comprises" par la machine, parce qu'elles sont destinées à s'intégrer dans un réseau sémantique, mais doivent également être lues directement par un être humain, car elles servent à la constitution d'outils bibliographiques imprimés. L'objectif de cette fiche n°2 est de distinguer les différents types d'information asso-

ciés au contenu de chaque document, en tenant compte des trois facteurs qui interviennent dans leur production (le documentaliste indexeur, le langage documentaire et la machine) et de leur double fonction (compréhension par la machine et lisibilité par l'être humain).

2 LES INFORMATIONS INDEXEES ET LEUR CLASSEMENT

En première approximation, on dira que les informations indexées d'un document sont constituées par l'ensemble des informations qui, au jugement de l'indexeur, représentent valablement le contenu de chaque document. En sont exclues toutes les autres informations concernant le document, telles que références bibliographiques, localisation du document, etc.

Les informations indexées comprennent, entre autres, les *structures lisibles*. Ce sont des informations exprimées en langue naturelle, et donc directement lisibles par un être humain, que la machine peut produire en sortie.

Les informations indexées sont susceptibles d'être classées à partir des trois critères qui suivent:

- (A) les procédés d'intervention de l'indexeur pour les produire
- (B) leurs fonctions
- (C) leur présentation.

2.1 PROCEDES D'INTERVENTION DE L'INDEXEUR

Le documentaliste peut produire certaines des informations indexées par *écriture directe*, c'est-à-dire en utilisant un ensemble de conventions qui vont lui permettre de créer une structure qui sera entrée en machine. Par exemple, le documentaliste peut produire la structure

Indexation t1 automatique al des textes t0.

à laquelle correspond la structure lisible

Indexation automatique des textes.

Par ailleurs, le documentaliste peut se servir d'une *instruction d'écriture*. Dans ce cas, il n'écrit pas directement l'information qu'il veut associer à un document, mais l'instruction pour que la machine crée cette information par *écriture médiate*.

Nous réservons l'expression *écriture par le documentaliste* pour désigner tout ce que le documentaliste indexeur écrit et associe au contenu d'un document, soit par écriture directe soit comme instruction d'écriture.

Ces conventions sont schématisées dans le tableau suivant:

ECriture PAR LE DOCUMENTALISTE	}	INSTRUCTIONS D'ECRITURE	}	INFORMATIONS INDEXEES
		INFORMATIONS INDEXEES PAR ECRITURE DIRECTE		
		INFORMATIONS INDEXEES PAR ECRITURE MEDIANE		

La composition des informations indexées par écriture directe sera traitée dans la fiche sur les *Structures engendrées par la grammaire de VI*.

2.1.1 LES INSTRUCTIONS D'ECRITURE

Toutes les instructions d'écriture déclenchent des opérations de la machine qui produisent des informations indexées. VI prévoit trois types d'instructions d'écriture:

- (A) Commutation
- (B) Permutation
- (C) Développement de l'anaphorique

2.1.1.1 L'INSTRUCTION DE COMMUTATION

Supposons que l'indexeur souhaite entrer en machine les structures qui correspondent aux structures lisibles suivantes:

1. *Indexation automatique des textes.*
2. *Traduction automatique des textes.*

Au lieu d'écrire directement

1. *Indexation t1 automatique al des textes t0.*
2. *Traduction t1 automatique al des textes t0.*

il peut produire la structure suivante par écriture directe:

1. *Indexation t1 automatique al des textes t0.*

1
2
3

(les numéros qui figurent au-dessous de la ligne sont ajoutés par la machine; cf. fiche sur l'Assistance de l'ordinateur à l'indexation) et ajouter l'instruction d'écriture

2. & l= traduction t1.

Cette instruction d'écriture fera produire à la machine, par écriture médiane, l'information indexée suivante:

2. *Traduction t1 automatique al des textes t0.*

Soit encore l'exemple suivant:

1. Gestion t1 automatisée a4 de la bibliothèque t0.

2-4 & 3= du centre de documentation t3, du prêt t1 interbibliothèque a1, de l'abonnement t1 aux périodiques t0.

Cette instruction d'écriture indique que la sous-structure
de la bibliothèque tO

doit être remplacée par les sous-structures

du centre de documentation t3

du prêt t1 interbibliothèque al

de l'abonnement t_1 aux périodiques t_0 .

Par application de cette instruction d'écriture, la machine produira, par écriture médiate, les structures suivantes:

2. Gestion t1 automatisée a4 du centre de documentation t3.
3. Gestion t1 automatisée a4 du prêt t1 interbibliothèque a1.
4. Gestion t1 automatisée a4 de l'abonnement t1 aux pério-

L'instruction de commutation provoque donc la substitution - dans un contexte invariant - d'une sous-structure X' à une autre sous-structure X appartenant à une structure complète Y . Comme résultat de cette substitution, on obtient une structure complète Y' , composée de l'élément X' et du contexte invariant.

Il est à remarquer que la structure Y subsiste en tant qu'information indexée par écriture directe. Les instructions d'écriture font donc produire à la machine des informations médiatees qui *s'ajoutent* à l'information directe et qui, mis à part le procédé d'obtention, présentent les mêmes caractères et accomplissent les mêmes fonctions que l'information indexée en écriture directe.

2.1.1.2 L'INSTRUCTION DE PERMUTATION

L'instruction de permutation fait modifier par la machine l'ordre des sous-structures dans une structure Y, de manière à obtenir la structure médiate Y'. Dans la structure médiate Y', on retrouve toutes les sous-structures de la structure Y, mais dans un ordre différent.

Exemple. Soit la structure:

Traitement tl par lots al en France al de l'information tl.

La structure lisible correspondante devrait être:

Traitement par lots en France de l'information.

Or, cette structure lisible est stylistiquement peu élégante et peu naturelle.

L'instruction d'écriture

(4,2,3)

ajoutée comme suit à l'information indexée:

Traitement t1 par lots al en France al de l'informa-
 1 2 3 4

tion t1 (4,2,3).

permet d'obtenir la structure lisible qui suit:

Traitement de l'information par lots en France.

L'instruction de permutation n'est destinée qu'à permettre une meilleure structure lisible; sa seule fonction concerne la sortie en langue naturelle. L'information indexée qu'elle crée est une structure de transition; on désigne par cette expression les informations indexées dont le rôle est uniquement de concourir à la formation de structures lisibles.

2.1.1.3 L'INSTRUCTION DE DEVELOPPEMENT DE L'ANAPHORIQUE

La grammaire de V1 prévoit la possibilité de l'anaphore. Soit la structure suivante et la structure lisible correspondante:

Centre de documentation t3 en Allemagne al sur l'urbanisme al.

Centre de documentation en Allemagne sur l'urbanisme.

Une autre structure peut reprendre anaphoriquement la première; par exemple:

Gestion t1 de ce centre t3.

Gestion de ce centre.

Mais il se peut que, dans un même document, il y ait deux centres de documentation qui soient mentionnés et que le document ne parle de la gestion que de l'un d'eux; par exemple, si l'on avait:

1. *Centre de documentation t3 en Allemagne al sur l'urbanisme al.*

2. *Centre de documentation t3 en France al sur l'urbanisme al.*
 et si on voulait ajouter la structure correspondant à l'information lisible

Gestion de ce centre de documentation en Allemagne sur l'urbanisme.

il faudrait produire la structure

3. *Gestion t1 de ce(1) centre t3.*

La référence, à l'intérieur des parenthèses, permet de lever l'ambiguïté et d'obtenir la structure lisible voulue.

Dans certains cas, il se peut que la reproduction, en structure lisible, de la totalité des compléments portant sur le terme auquel renvoie l'anaphorique ne soit pas naturelle, sur le plan stylistique. Dans l'exemple précédent,

on peut préférer, comme structure lisible, l'expression

Gestion de ce centre de documentation allemand.

Dans ce cas, l'indexeur ne doit pas produire la structure (3) mais la structure (3 bis), où il a ajouté une instruction de développement de l'anaphorique, sous la forme suivante, entre crochets carrés:

3 bis. *Gestion t1 de ce(1) centre [de documentation allemand] t3.*

De même que l'instruction de permutation, l'instruction de développement de l'anaphorique n'est destinée qu'à obtenir une meilleure lisibilité de la sortie en langue naturelle. Elle crée une structure de transition.

2.2 LES FONCTIONS DES INFORMATIONS INDEXEES

Il a été rappelé au début de cette fiche que les informations indexées, associées au contenu de chaque document, doivent à la fois être comprises par la machine et être lisibles par un être humain; cela nous permet de distinguer deux fonctions: fonction de *compréhension* (sous-entendu, par la machine) et fonction de *lisibilité* (sous-entendu, par un être humain). Les informations indexées peuvent se classer selon qu'elles sont destinées à assurer l'une ou l'autre de ces fonctions ou les deux à la fois.

L'indexeur, soit par écriture directe, soit par le biais d'une instruction de commutation, produit une structure, qui sera désormais appelée *structure grammaticale*. La structure suivante est donc une structure grammaticale:

Edition t1 par microfilm al de documents t0 en anglais al.

Dans cette structure grammaticale, nous distinguons deux parties: l'*information de base*, d'une part, et les *articulations*, d'autre part. L'information de base est la partie d'une structure grammaticale qui assure l'essentiel de la fonction de compréhension et participe, avec les articulations, à la fonction de lisibilité. Les articulations assurent, avec l'information de base, la fonction de lisibilité et peuvent, dans certaines conditions, contribuer à la fonction de compréhension.

Dans une structure grammaticale, l'information de base est constituée par les procédés, représentés par des indicateurs (*t1*, *al*, etc.), qui servent également de bornes entre les sous-structures, et par un certain nombre de caractères de la sous-structure qui forment le *radical discriminant* de la notion véhiculée par la sous-structure. On appelle ainsi, par convention, la séquence minimale de caractères alpha-numériques (et éventuellement de blancs) nécessaire et suffisante pour que la machine reconnaisse la notion en la distinguant sans ambiguïté des autres notions présentes dans le lexique. Le radical discriminant de la notion est fixé au moment de la constitution du lexique a-

vec le triple souci d'obtenir (a) la forme la plus économique pour le traitement, (b) une discrimination dénuée de toute ambiguïté, (c) une parfaite superposition avec les caractères de la sous-structure exprimée en langue naturelle (ce qui assure la transparence de ce radical discriminant pour l'indexeur). Ces trois objectifs sont en effet conciliables dans le cadre d'un système qui, comme VI, prévoit un lexique fermé. En règle générale, la machine décode, par un jeu de troncatures et éventuellement de masques, le dernier mot orthographique qui précède l'indicateur de procédé. L'indexeur, qui jouit d'une très grande liberté de rédaction (en fonction de la structure lisible qu'il souhaite obtenir) pour le contexte qui précède, à l'intérieur de la sous-structure, ce dernier mot, appelé *support de radical*, doit respecter pour le support de radical des contraintes plus rigoureuses, mais qui ne vont pas à l'encontre des habitudes les plus naturelles. Dans des cas plus rares, la machine a des instructions pour repérer de l'information de base en amont du support de radical.

On peut constater que:

(A) Des informations telles que les marques de genre et de nombre et, la plupart du temps, les morphèmes de dérivation, ne font pas partie de l'information de base, mais des articulations.

(B) La distinction entre *information de base* et *articulations*, qui est essentielle pour comprendre le fonctionnement du système, est presque totalement transparente pour l'indexeur, si l'on excepte les indicateurs de procédé (*tl*, *al*, etc.), qui sont nettement détachés. L'indexeur n'a pas à indiquer matériellement cette distinction et, à quelques détails près, il n'a pas à s'en soucier.

L'information de base de la structure grammaticale déjà citée:

Edition tl par microfilm al de documents to en anglais al.

est la suivante (les radicaux discriminants sont représentés ici par des capitales, les articulations sont soulignées):

EDITION tl par MIcroFILm al de DOCuments to en ANGLais al.

Les structures grammaticales ne peuvent être produites que par écriture directe ou par une instruction de commutation; en revanche, les instructions de permutation et de développement de l'anaphorique, qui n'assurent que la fonction de lisibilité, ne créent pas de structure grammaticale ni aucune information de base nouvelle (elles créent des structures de transition, cf. §2.1.1.2). On voit qu'une structure grammaticale se caractérise par le fait de présenter une information de base spécifique, qui n'est identique dans aucune autre struc-

ture grammaticale du même document.

2.3 LA PRESENTATION DES INFORMATIONS INDEXEES

Les informations indexées sont susceptibles d'apparaître sous deux formes, soit comme *structures*, soit comme *classement de structures*.

On n'a fait état, jusqu'ici, que des informations indexées présentées sous la forme de structures: structures grammaticales, composées d'information de base et d'articulations, structures de transition et structures lisibles. Or, VI ajoute aux informations indexées qui prennent la forme de structures grammaticales (ou de structures lisibles) une autre information: le classement de ces structures selon leur pertinence. Ce classement admet deux possibilités:

expressions documentaires (= ED)

sous-expressions documentaires (= SED).

Une structure grammaticale (et la structure lisible correspondante) est une ED si elle justifie, à elle seule, que le document qu'elle représente "tombe" au cours d'une interrogation. Une structure grammaticale (et la structure lisible correspondante) est une SED si elle ne justifie pas, à elle seule, que le document qu'elle contribue à représenter "tombe".

Le classement en ED ou SED s'applique aussi bien aux structures créées à la suite d'une instruction de commutation qu'à celles qui sont produites par écriture directe. Toute structure grammaticale est soit une ED, soit une SED. Toute SED s'intègre dans une ED. Les structures lisibles portent automatiquement le classement en ED ou en SED des structures grammaticales dont elles dérivent. Les ED et les SED sont ouvertes et automatiquement numérotées par l'ordinateur en réponse à une commande simple de l'indexeur (cf. *fiche Assistance de l'ordinateur à l'indexation*).

Exemple: Soit le document A. Ce document décrit la bibliothèque nationale du Mexique, le centre de documentation en éducation du même pays et la coopération qui existe entre ces deux organismes. Chacun de ces trois thèmes est traité de telle manière qu'il est justifié que le document tombe en fonction de l'un quelconque d'entre eux. Les structures grammaticales correspondantes seront, toutes les trois, des ED. Le document A sera donc indexé comme suit:

1. *Bibliothèque t0 nationale al au Mexique al.*
2. *Centre de documentation t3 en éducation al au Mexique al.*
3. *Cette bibliothèque t0 en coopération avec r ce centre t3.*

Soit le document B. Ce document décrit la bibliothèque nationale du Mexique et parle de la coopération de cette bibliothèque avec le centre de documentation en éducation du même pays, mais ne décrit pas ce dernier organisme,

qui n'intervient dans le document qu'au titre de ses relations avec la bibliothèque nationale. La structure grammaticale qui correspond au centre de documentation ne sera, par conséquent, qu'une SED. Le document B sera donc indexé comme suit:

1. *Bibliothèque t0 nationale al au Mexique al.*

2.1 *Centre de documentation t3 en éducation al au Mexique al.*

2 Cette bibliothèque t0 en coopération avec r ce centre t3.

Le classement des structures grammaticales, librement effectué par l'indexeur en fonction de sa compréhension du document, est une information importante dont la machine doit se servir lors de la recherche documentaire et de la constitution du réseau sémantique: à ce titre, l'information de classement concourt à la fonction de compréhension.

3. LES STRUCTURES LISIBLES

Les structures lisibles constituent un des types d'informations indexées; elles permettent la sortie en langue naturelle produite par la machine selon différents formats adaptés aux divers produits documentaires que l'on souhaite obtenir (bulletins de résumés, index permuté, diffusion sur profil, etc.) ainsi qu'à un affichage lisible sur l'écran d'un terminal. On aura soin de distinguer les *structures lisibles*, qui sont des informations indexées, produites par la machine à partir des structures grammaticales, de leur classement et des structures de transition et donc, en dernière analyse, associées à un document particulier par des choix de l'indexeur (même s'il n'écrit jamais directement les structures lisibles) et les multiples *formats de sortie* possibles, qui ne sont que des conversions automatiques des structures lisibles selon des procédures générales, dépendant non du contenu des documents mais du type de sortie souhaité.

4 RECAPITULATION

Les informations indexées ont été classées, dans les pages précédentes, selon trois critères (procédés d'intervention, fonction et présentation), chacun de ces critères permettant de sous-catégoriser les informations indexées en plusieurs types différents. Ces critères peuvent se combiner entre eux, d'où il résulte le tableau récapitulatif de la page suivante.

Le tableau récapitulatif permet de caractériser, par énumération, la notion d'information indexée, complétant ainsi la première approximation présentée au § 2.

				Informations indexées				
				Classe- ment de structures	Structures			
					Str. grammatic.		Str. de transi- tion	Str. lisibles
					Inf. base	articul.		
Types de production des infor- mations indexé- es	Procé- dés d'intér- vention de l'in- dexeur	Ecriture directe		+	+	+	-	-
		Ecriture médiante (Ins. d'écri- ture)	Ins. de commu- tation	+	+	+	-	-
			Ins. de permu- tation	-	-	-	+	-
			Ins. de dév. anaphor.	-	-	-	+	-
		Production par la machi- ne à partir des str.gram., de leur clas., et des str. de transition		-	-	-	-	+

TABLEAU RECAPITULATIF

(Le symbole "+" indique que le type d'information indexée indiqué en haut de la colonne est susceptible d'être obtenu par le type de production indiqué à gauche de la ligne.)



Fiche n°3 (décembre 1979)

G.G.Bès et P.-M.Fauchère

Assistance de l'ordinateur à l'indexation. Principes généraux

Le système Vercingétorix prévoit une assistance de l'ordinateur dans le cadre d'un dialogue interactif entre l'homme et la machine, aussi bien au moment de l'indexation qu'au stade de la recherche documentaire proprement dite. Dans ce système, l'ordinateur n'est pas seulement un outil de stockage des données et de tri des références pertinentes par application de quelques opérations logiques simples. Il doit guider le travail de l'utilisateur, en vérifiant la cohérence et en lui montrant les conséquences de ses choix, en lui proposant des stratégies d'indexation et de recherche, ce qui suppose, dans ce dernier cas surtout, la mise en oeuvre de procédures complexes, pouvant faire appel aux possibilités de l'intelligence artificielle et de l'auto-apprentissage de la machine. La structure même du langage de représentation du contenu des documents est commandée par cette interaction entre l'homme et la machine (voir, dans la fiche n°2: *Les informations indexées*, ce qui est dit, au § 2.2, sur les fonctions de *compréhension* et de *lisibilité*).

Concrètement, le documentaliste se trouvera, pour dialoguer avec l'ordinateur, devant un ensemble clavier + écran.

L'aide de l'ordinateur à l'indexation peut intervenir à différents niveaux (on n'envisagera ici que les opérations qui se situent après l'entrée dans la zone d'indexation, c'est-à-dire celles qui concernent la représentation du contenu des documents):

(1) Procédures d'assistance matérielle.

L'ordinateur épargne à l'indexeur un certain nombre de tâches matérielles: c'est notamment le système de numérotation automatique des ED, des SED et des items, et de renumérotation automatique lorsqu'il y a eu des

modifications en cours d'indexation.

Afin de faciliter la lecture des exemples de la fiche n°4, on donne ci-dessous quelques indications plus précises sur la numérotation automatique des ED, des SED et des items.

Lorsque l'indexeur, par une commande spécifique, a indiqué l'entrée dans la zone d'indexation, la machine ouvre automatiquement la première ED; elle fait apparaître sur l'écran le chiffre 1. L'indexeur écrit alors la première suite d'items constituant une structure grammaticale. S'il s'agit d'une ED, l'indexeur frappe la touche RETURN; le chiffre 2, numéro de la deuxième ED, s'inscrit alors sur l'écran au début de la ligne suivante. Si, au contraire, il s'agit d'une SED, l'indexeur frappe un point et le chiffre 2, numéro de la deuxième SED appartenant à la première ED, s'inscrit sur la même ligne. Dans les deux cas, l'indexeur peut alors entrer une deuxième structure grammaticale, et ainsi de suite.

La numérotation automatique des items n'est utile que pour permettre l'introduction des instructions d'écriture de commutation ou de permutation (et très exceptionnellement pour certains cas d'utilisation de l'anaphore). Elle épargne à l'indexeur la peine d'écrire à nouveau les items concernés. Cette numérotation, dans le cas des instructions de commutation ou de permutation, est déclenchée par le signe qui indique l'ouverture de l'instruction et affecte, sauf indication contraire, la structure grammaticale qui précède immédiatement.

(2) Procédures de contrôle

L'ordinateur contrôle la correction des structures grammaticales, au fur et à mesure qu'elles sont écrites par l'indexeur, et vérifie la cohérence d'ensemble de chaque indexation. Il envoie éventuellement des messages d'erreur. Il signale ainsi toute entorse aux règles de la syntaxe: il intervient si, par exemple, l'indexeur a introduit une succession interdite de procédés, s'il a utilisé irrégulièrement l'anaphorique, s'il a violé des contraintes sémantiques. Il repère de même les instructions d'écriture incorrectes.

Dans les cas précédemment cités, l'ordinateur peut détecter à coup sûr des éléments inacceptables en fonction des règles qu'on lui a fournies. Mais l'extension de ce contrôle reste limitée par le fait que le langage documentaire a éliminé une grande partie de la redondance, ce qui empêche souvent les contrôles croisés. Chaque fois, et c'est le cas le plus fréquent, que l'in-

dexeur se trouve devant un choix libre, la machine n'a pas les moyens d'examiner la validité de la solution choisie, et si l'indexeur commet une étourderie, la machine ne pourra la repérer directement. Il est donc prévu un contrôle plus élaboré, de caractère statistique, portant sur la cohérence des éléments syntaxiques et des éléments lexicaux d'une part, sur celle des articulations et de l'information de base d'autre part. Chaque fois que l'indexation proposée enfreindra les données de certaines tables, prévues dès le départ ou calculées par la machine à partir du stock des indexations déjà mémorisées, la machine alertera l'indexeur, qui restera maître de la décision finale.

(3) Procédures d'optimisation de l'indexation

Pour construire les structures grammaticales, l'indexeur doit avoir accès au lexique et aux listes qu'il contient (notions, procédés acceptés par chaque notion, traits sémantiques associés aux items). S'il veut relier deux phrases documentaires dans une construction documentaire, il doit connaître la liste des opérateurs r qui peuvent intervenir. Il peut consulter ces listes dans le Manuel d'acquisition et d'utilisation, mais il peut également les faire apparaître à l'écran.

Il est souhaitable que l'indexeur puisse se référer à des indexations antérieurement produites. Il doit donc avoir accès à la mémoire documentaire et aux ED qu'elle contient. Il pourra consulter aussi les tables statistiques calculées par la machine à partir des indexations stockées ou des recherches déjà effectuées.

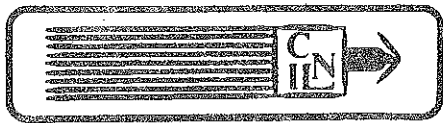
Lorsqu'une ED (ou une SED) a été formée, il est important que l'indexeur puisse vérifier, commodément et sans avoir à se livrer à une gymnastique intellectuelle complexe, les répercussions de ce qu'il a écrit. L'ordinateur doit donc fournir à sa demande les sorties en langue naturelle dans les différents formats souhaités. Il doit surtout permettre à l'indexeur de connaître exactement la situation de l'ED dans le réseau sémantique, c'est-à-dire dans l'ensemble des relations déterminées par l'organisation syntaxique de l'ED et par les traits sémantiques associés aux items; c'est ce réseau sémantique qui, dans la phase de recherche documentaire, commande la possibilité, à partir d'une question donnée, d'élargir, de restreindre ou de déplacer automatiquement le champ de la recherche. Il est donc particulièrement utile que l'indexeur puisse simuler des procédures d'interrogation pour voir dans quelles conditions le document tomberait à la suite de modifications progressives et en partie

automatiques du champ de la recherche. A la limite, compte tenu des résultats que l'indexeur veut obtenir et de ceux qu'il veut éviter, l'ordinateur pourra calculer la situation optimale dans le réseau sémantique et proposer, en conséquence, à l'indexeur la formulation correspondante de l'ED.

(4) Procédures d'aide à l'acquisition

Le système doit prévoir enfin une procédure qui guidera pas à pas l'indexeur débutant dans sa démarche. Cette procédure, évidemment, alourdira et ralentira les opérations d'indexation. Aussi pourra-t-elle être partiellement ou totalement déconnectée par l'indexeur plus expérimenté et, éventuellement, rappelée en cas d'hésitation.

Pourvu de ces procédures d'assistance, le système doit permettre une utilisation simple et efficace du langage documentaire pour représenter le contenu des documents. Il doit être accessible, après une période de formation spécialisée dont la durée ne soit pas excessive, à des documentalistes capables par ailleurs de comprendre et d'interpréter correctement les documents à analyser.



Fiche n°4 (août 1979)

G.G. Bès et P.-M. Fauchère

Exemples d'indexation (Série n°1)

On trouvera dans cette série n°1 des exemples d'indexation produits à l'aide du langage documentaire de V1, tel que ce langage est défini en août 1979. Les dix exemples de cette série ont tous été présentés suivant le même modèle.

Il faut remarquer que:

(1) Le résumé en langue naturelle (zone 3 de chaque exemple) a été rédigé en fonction du résumé original (zone 1, en haut à droite) et de l'indexation actuelle par descripteurs (zone 2), qui s'appuie elle-même sur le Thésaurus Sciences de l'Information (Paris, C.N.R.S., 1977). Ces deux éléments ont été empruntés au Bulletin signalétique, section 101. Les contraintes générales sur le résumé en langue naturelle sont donc les suivantes:

(a) le contenu du résumé publié dans le B.S. est accepté tel qu'il y apparaît, sans être contrôlé par rapport au document original;

(b) l'indexation figurant dans le B.S. est utilisée pour éclairer ce contenu; on a cherché à rendre tous les concepts exprimés par les descripteurs, sans nécessairement les reprendre individuellement;

(c) à partir de l'image du document que fournissent le résumé et l'indexation du B.S., on s'est donné pour objectif de faire apparaître tout ce qui justifiait la sortie du document dans la phase de recherche documentaire, mais en se limitant aux notions prévues par le lexique de V1.

La fonction de V1 n'est pas en principe d'assurer la compréhension du document original; la qualité de ce résumé en langue

naturelle ne peut donc pas être considérée comme une manière de tester V1. Plutôt, pour vérifier V1, on se donne un résumé en langue naturelle, que l'on veut le plus proche possible du contenu du texte original; la vérification de ce résumé rentre dans le cadre de la problématique et de la méthodologie de la contraction de texte (cf. Bellot et al., La contraction de texte).

(2) Le lexique de V1 n'a pas été entièrement établi. Or, les indexations des exemples qui suivent supposent l'existence d'un lexique déterminé; il se peut, par conséquent, que certaines décisions concernant des items lexicaux, dans des cas particuliers, soient révisées plus tard en fonction des conditions d'adéquation internes à l'organisation de ce lexique. Mais tous les exemples présentés supposent fixées les contraintes générales d'organisation du lexique (les procédés, les traits sémantiques apportés par les procédés, etc.) qui, elles, ne peuvent être modifiées de manière ad hoc.

(3) Dans la zone 4 ("Indexation avec V1: écriture par le documentaliste"), on trouve ce que le documentaliste opérant avec V1 doit écrire et des indications fournies par la machine dans le cadre de sa fonction d'assistance à l'indexation. Les apports de la machine sont ici la numérotation automatique des ED (= expressions documentaires) et des SED (= sous-expressions documentaires) ainsi que la numérotation des items d'une structure grammaticale, numérotation produite à la demande du documentaliste pour permettre des renvois; v. fiche n°3: Assistance de l'ordinateur à l'indexation, et ci-dessous: Observations sur la zone 4; conventions de notation.

(4) La zone 5 illustre un des types possibles de sortie en langue naturelle (adapté aux bulletins de résumés), tels que la machine est susceptible de les produire dans son état actuel; la qualité de ces textes produits automatiquement par la machine à partir de l'information apportée par l'indexation (cf.(3) ci-dessus) pourra être notablement améliorée.

OBSERVATIONS SUR LA ZONE 4 ; CONVENTIONS DE NOTATION

Pour faciliter la lecture des exemples qui suivent, on trouvera ci-dessous les conventions de notation, accompagnées des éléments sur l'organisation de la grammaire de V1 qui sont nécessaires pour la compréhension de ces exemples. Il ne s'agit que d'indications partielles; on abordera le problème plus en détail dans la fiche (à paraître) Structures engendrées par la grammaire de V1.

La grammaire de V1 est constituée d'un lexique et d'un ensemble de règles syntaxiques. Elle spécifie les structures grammaticales.

Le lexique est composé d'un ensemble fini de notions. Ces notions sont catégorisées en fonction de leur aptitude à acquérir un ou plusieurs procédés, parmi une liste finie et très restreinte. On trouve indiqués, dans le lexique, les procédés qui peuvent être acquis par chaque notion. Une notion qui a acquis un procédé déterminé et qui apparaît dans une structure grammaticale constitue un item; celui-ci est donc composé d'une notion suivie de l'indicateur de procédé (t0, a1, etc.); il est précédé, le cas échéant, d'une articulation (cf. fiche n°2).

Les procédés ont une double fonction: (a) indiquer la relation syntaxique d'une notion dans une structure; (b) ajouter, dans certains cas, des traits sémantiques à cette notion. Il existe deux catégories de procédés: T et A. En règle générale, les procédés T correspondent aux items qui constituent les termes ou "noyaux" d'une organisation syntaxique; les procédés A correspondent aux items qui sont des attributs des T.

La catégorie T comprend: t0 ou t1, t2, t3.

- t0 est utilisé pour les notions qui, en langue naturelle, sont exprimées par un nom ou par une construction nominale (Ex.: bibliothèque t0, document t0, recherche scientifique t0).

- t1 est utilisé quand on part, comme notion de base, d'un verbe en langue naturelle (Ex.: financement t1, diffusion t1, accès t1, condensation t1). Le trait sémantique qu'il introduit est "action et/ou effet de". t0 et t1 sont exclusifs; une notion, dans le lexique de V1, admet soit t0, soit t1.

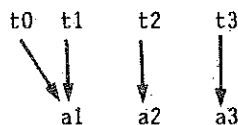
- t2 introduit le trait sémantique "machine qui sert à"; la notion de base correspondante, en langue naturelle, est un verbe (Ex.: agrandisseur t2).

- Le trait sémantique introduit par t3 dépend du comportement de la notion respective à l'égard des procédés t0 et t1. Si la notion admet t0, le trait introduit par t3 sera "personne dont l'activité est caractéristique de" (Ex.: bibliothécaire t3). Si la notion admet t1, le trait sera "personne ou organisme qui" (Ex.: indexeur

t3, centre de documentation t3). L'ambiguïté entre "personne" et "organisme" est levée soit par des suffixes (par exemple, -eur dans indexeur t3, traducteur t3), soit par des expressions en langue naturelle telles que centre de, service de, organisme de, qui doivent précéder la notion accompagnée de l'indicateur de procédé (Ex.: service de documentation t3, organisme de transmission t3).

La catégorie A comprend: a1, a2, a3, a4.

- a1, a2, a3 introduisent le même trait sémantique, très large, de "propre ou relatif à (la notion de), en relation avec (la notion de)". La relation entre les procédés T et les procédés A s'établit comme suit:



a1 est le procédé qui fait acquérir la fonction d'attribut aux items en t0 et t1; a2 fait de même avec t2 et a3 avec t3.

- le procédé a4 introduit le trait sémantique "ayant subi l'action de". Il ne s'utilise qu'avec les notions qui admettent t1.

Le système de procédés, tel qu'il apparaît ici, conserve volontairement une certaine part de redondance. Dans la fiche sur les Structures engendrées par la grammaire de V1, on examinera dans quelle mesure il est opportun ou non de limiter cette redondance.

Les structures grammaticales spécifiées par V1 sont constituées par un enchaînement d'items (avec, le cas échéant, leurs articulations), auquel les relations syntaxiques entre les items, sous la forme de structures hiérarchiques emboîtées, sont automatiquement associées par la machine.

Sur le plan documentaire, les structures grammaticales sont classées en ED et en SED (cf. fiche n°2); elles sont numérotées en tant que telles.

Sur le plan grammatical, les structures grammaticales engendrées par V1 sont de deux types: les phrases documentaires et les constructions documentaires. Les phrases documentaires sont exclusivement constituées d'items dont les procédés appartiennent à une des catégories indiquées ci-dessus ($T = \{ \underline{t0}, \underline{t1}, \underline{t2}, \underline{t3} \}$; et $A = \{ \underline{a1}, \underline{a2}, \underline{a3}, \underline{a4} \}$). Les constructions documentaires sont des structures grammaticales composées de deux phrases documentaires mises en relation par un prédicat. Ce prédicat appartient à une liste finie et restreinte; dans les exemples qui suivent, on trouve notamment avec l'objectif de, au moyen de, influençant; ces prédicats sont suivis du symbole "r", qui joue un rôle analogue à celui des indicateurs de procédé.

Exemples:

Phrases documentaires

Bibliothèque t0 universitaire a1 au Mexique a1.

Centre de documentation t3 au Canada a1.

Constructions documentaires

Cette bibliothèque t0 en coopération avec r ce centre t3.

Dans une construction documentaire, r et le prédicat qui précède sont traités comme un item, qui pourra être numéroté comme les autres items et qui sera susceptible d'être repris par l'anaphore.

Les relations syntaxiques entre les items d'une structure grammaticale sont déduites par la machine à partir des indicateurs de procédé et de l'emplacement des items. Dans certains cas, cependant, apparaît un symbole supplémentaire, le symbole "b", qui note une relation syntaxique particulière: les items de la catégorie A qui suivent b sont en relation avec l'item de la catégorie A qui précède immédiatement b (et non avec le premier item de catégorie I qui se trouve plus à gauche). Par exemple, les structures grammaticales

(a) Catalogue t0 de documents a1 b en français a1.

(b) Catalogue t0 de documents a1 en français a1.

signifient respectivement:

(a') Catalogue de documents qui sont en langue française.

(b') Catalogue en français de documents.

Les sous-structures qui renvoient à une phrase documentaire antérieure (par anaphore) se présentent de la manière suivante:

Forme anaphorique	Référence	Terme repris	Développement de l'anaphorique	Indicateur de procédé
1	2	3	4	5

(Les éléments 1, 3 et 5 sont obligatoires; les éléments 2 et 4, optionnels. Peuvent s'y ajouter, devant la forme anaphorique, des éléments d'articulation.)

Les formes anaphoriques sont celles des démonstratifs ce, cet, cette, ces.

Le terme repris est le calque d'un antécédent qui est apparu antérieurement dans une autre phrase documentaire correspondant au même document. Il peut s'agir selon les cas: a) de la notion présente dans un item; b) de cette notion précédée d'une expression en langue naturelle propre à lever l'ambiguïté d'un t3; c) de la notion présente dans le prédicat d'une construction documentaire.

Exemples:

(a) Bibliothèque t0 universitaire a1.

Financement t1 de cette bibliothèque t0.

(b) Centre de documentation t3.

Financement t1 de ce centre t3.

(c) Ce centre t3 en coopération avec r cette bibliothèque t0.

Cette coopération t1 en France a1.

L'indicateur de procédé est celui qui est exigé dans la structure grammaticale où apparaît l'item anaphorique. Exemple:

Bibliothèque t0 universitaire a1.

Catalogue t0 de cette bibliothèque a1.

La référence se rencontre lorsqu'il peut y avoir ambiguïté sur l'identification de l'antécédent du terme repris. Elle se note, entre parenthèses, immédiatement après la forme anaphorique et comprend, dans cet ordre, le numéro de l'ED et, si nécessaire, ceux de la SED et de l'item (Si, dans les informations indexées qui précèdent, il n'y a qu'un seul terme susceptible d'être l'antécédent, la référence est inutile; de même, si dans l'ED il n'y a qu'un seul terme qui puisse être l'antécédent, les numéros de SED et d'item ne sont pas nécessaires; et le numéro d'item n'est indiqué que s'il y a ambiguïté à l'intérieur d'une SED).

Le développement de l'anaphorique est indiqué entre crochets droits. Pour l'utilisation de cette instruction d'écriture, se reporter à la fiche n°2.

Indications complémentaires

(a) Numérotation:

Elle porte sur les ED (avec passage à la ligne pour chaque nouvelle ED), sur les SED (sans passage à la ligne à l'intérieur d'une même ED) et sur les items (numérotation en-dessous des items, chaque fois que c'est nécessaire pour permettre d'écrire une instruction de commutation ou de permutation ou une référence anaphorique).

(b) Sous-structure:

Aucune différence de notation entre articulations, expressions en langue naturelle destinées à lever l'ambiguïté de certains procédés, radicaux discriminants des notions.

(c) Instruction de commutation:

On trouve successivement: (1) le numéro de l'ED ou de la SED à créer; (2) le signe "&"; (3) le(s) numéro(s) de la sous-structure ou des sous-structures à remplacer dans la structure grammaticale précédente; s'il y en a plusieurs, elles sont nécessairement contiguës; (4) le signe "="; (5) la ou les sous-structures de remplacement.

Le nombre de sous-structures n'est pas nécessairement le même de chaque côté du signe "=". L'ensemble de sous-structures figurant en (5) vient, en bloc, prendre la place de l'ensemble de sous-structures mentionné en (3).

La même instruction de commutation peut permettre la création de plusieurs structures grammaticales nouvelles. Dans ce cas, à la sous-structure ou à l'ensemble de sous-structures indiqué en (3) correspondent en (5) plusieurs substituts (composés chacun d'une ou de plusieurs sous-structures), délimités par des virgules.

Un élément d'articulation peut éventuellement s'insérer entre le dernier indicateur de procédé d'un substitut et le point ou la virgule qui marque la limite de ce substitut.

Ces différentes possibilités sont illustrées par les exemples 3 (ED n°4-11) et 10 (SED n°1.3).

(d) Instruction de permutation:

Elle se place à la suite de la structure grammaticale sur laquelle elle porte. On note entre parenthèses, en utilisant la numérotation des items, les modifications de l'ordre des sous-structures que l'on désire obtenir en structures lisibles.

EXEMPLE 1

(1) B.S. 101, 1978, n°90.

78-101-90. SEETHARAMA (S.). (Indian stat. inst., Bangalore 560003). *Library and information science research in India: An analytical study based on the input to ISORID. (Developments in library science. 13).*

Library Sci. Slant Document., India, (1976), 13, no 3-4, 121-137. Description du programme ISORID de l'UNESCO, et de la recherche en bibliothéconomie et science de l'information en Inde. On identifie les projets, les organismes engagés dans ces activités, le financement, la durée des projets, les moyens de diffusion des activités de recherche.

(2) Indexation actuelle par descripteurs

Recherche scientifique, Bibliothéconomie, Science information, Projet, Organisme recherche, Financement, Diffusion information, Inde.

(3) Résumé en langue naturelle

Recherche en bibliothéconomie et science de l'information en Inde: organismes, projets, financement, diffusion.

(4) Indexation avec V1: écriture par le documentaliste

1. Recherche scientifique t0 en Inde a1, en bibliothéconomie a1 et science de l'information a1.

2. Organismes t0 pour cette recherche a1.

1 2

3. & 1= projets t0.

4. Financement t1 de cette recherche t0.

1 2

5. & 1= diffusion t1.

(5) Sortie en langue naturelle produite par la machine

1. Recherche scientifique en Inde, en bibliothéconomie et science de l'information.

2. Organismes pour cette recherche.

3. Projets pour cette recherche.

4. Financement de cette recherche.

5. Diffusion de cette recherche.

EXEMPLE 2

- (1) B.S. 101, 1978, n°1364.

78-101-1364. PALMER (J. W.). (California state univ., Fullerton, Calif. 92634). The British Film Institute. Resources, publications, services.

Spec. Librar., U.S.A., (1977), 68, no 9, 313-319, bibl. (62 réf.).

On décrit le British Film Institute qui est un des centres les plus importants dans le monde pour les films scolaires et qui constitue un modèle dans ce domaine. Ses trois départements principaux sont présentés: la division information et documentation, le département des publications et celui des archives nationales du film; leurs activités sont passées en revue. On insiste sur le fait que le rayonnement de l'Institut est mondial grâce à ses publications et à son service de réponse aux questions. (INTD).

- (2) Indexation actuelle par descripteurs

Centre documentation, Document cinématographique, Education, Royaume-Uni.

- (3) Résumé en langue naturelle

Centre de documentation et filmothèque au Royaume-Uni pour les films éducatifs.

- (4) Indexation avec V1: écriture par le documentaliste

1. Centre de documentation t3 au Royaume-Uni a1 pour les films a1 b

1
2
3

éducatifs a1.

4
2. & 1= filmothèque t0.

- (5) Sortie en langue naturelle produite par la machine

1. Centre de documentation au Royaume-Uni pour les films éducatifs.
2. Filmothèque au Royaume-Uni pour les films éducatifs.

EXEMPLE 3

- (1) B.S., 101, 1979, n°1704.

79-101-1704. Die Universitätsbibliothek Saarbrücken. Bericht über die Jahre 1973 und 1974 mit einer historischen Einleitung anlässlich des 25-jährigen Bestehens der Bibliothek (La bibliothèque universitaire de Saarebruck. Rapport d'activité sur les années 1973 et 1974, avec une introduction historique à l'occasion du 25ème anniversaire de la bibliothèque). VINZENT (Oswin)

DEU, Saarbrücken: Universitätsbibl., (1975), 89 p., 21 cm

Faisant suite à l'historique de la bibliothèque, le rapport d'activité, présenté sous la forme d'un texte et de tableaux statistiques, comporte essentiellement les chapitres suivants: 1) l'accroissement du fonds et les acquisitions, réparties par facultés, par disciplines, par types de documents, par moyen d'acquisition (achats, échanges, dons), avec le montant du budget utilisé et les origines de financement; 2) les catalogues (notamment accroissement des unités entrées); 3) l'activité d'information; 4) les salles de lecture et de travail, leur fonds (documents primaires, catalogues, ouvrages de référence); 5) l'utilisation, les utilisateurs, le prêt sur place et le prêt interbibliothèque; 6) les services de photographie, reprographie et imprimerie (équipements et produits); 7) l'activité de formation professionnelle; 8) la structure de la gestion. On rapporte enfin l'histoire, le développement, et les activités du département médecine de la bibliothèque.

- (2) Indexation actuelle par descripteurs

Bibliothèque universitaire, Historique, Rapport activité, Statistiques, Acquisition, Catalogue, Fonds, Utilisation, Prêt, Reprographie, Formation professionnelle, Gestion, Budget, Bibliothèque spécialisée, Médecine, Allemagne (république fédérale).

- (3) Résumé en langue naturelle

La bibliothèque universitaire de Sarrebruck. Historique. Rapport d'activité et statistiques concernant les acquisitions, les catalogues, les fonds, l'utilisation, les activités de prêt, de reprographie, de formation professionnelle, la gestion, le budget. Présentation de la section médicale de la bibliothèque.

- (4) Indexation avec V1: écriture par le documentaliste

1. Bibliothèque t0 universitaire a1 en Allemagne (république fédérale) a1.

2. Historique t0 de cette bibliothèque a1.

3. Statistiques t0 et rapport t0 d'activité a1 concernant les
1 2 3
acquisitions a1 b de cette bibliothèque a1.

4 5
4-11. & 4= les catalogues a1 b, les fonds a1 b, l'utilisation a1 b, la gestion a1 b, le budget a1 b, le prêt a1 b dans, la reprographie a1 b dans, la formation a1 b professionnelle a1 dans.

12. Section t0 médicale a1 de cette bibliothèque a1.

- (5) Sortie en langue naturelle produite par la machine

1. Bibliothèque universitaire en Allemagne (république fédérale).

2. Historique de cette bibliothèque.
3. Statistiques et rapport d'activité concernant les acquisitions de cette bibliothèque.
4. Statistiques et rapport d'activité concernant les catalogues de cette bibliothèque.
5. Statistiques et rapport d'activité concernant les fonds de cette bibliothèque.
6. Statistiques et rapport d'activité concernant l'utilisation de cette bibliothèque.
7. Statistiques et rapport d'activité concernant la gestion de cette bibliothèque.
8. Statistiques et rapport d'activité concernant le budget de cette bibliothèque.
9. Statistiques et rapport d'activité concernant le prêt dans cette bibliothèque.
10. Statistiques et rapport d'activité concernant la reprographie dans cette bibliothèque.
11. Statistiques et rapport d'activité concernant la formation professionnelle dans cette bibliothèque.
12. Section médicale de cette bibliothèque.

EXEMPLE 4

(1) B.S. 101, 1978, n°389.

78-101-389. PELOU (P.). Les développements de la recherche documentaire automatisée dans les bibliothèques universitaires.

Bull. Dica, Fr., (1977), 2, no 5, 29-34, (Nouv. accès inf. bibl. téléinf. Journ.; 1977).

Activités de la DICA (Division de la Coopération et de l'Automatisation) en matière de recherche documentaire automatisée, dont l'importance est soulignée. Un programme pilote, réalisé à la bibliothèque du centre universitaire d'Orsay, a permis d'étendre cette expérience à d'autres bibliothèques et de leur permettre l'accès en conversationnel à diverses bases de données. Les problèmes de formation du personnel pour ces techniques nouvelles sont posés. Trois unités de travail ont été développées pour concourir à l'organisation du service de recherche: une bibliothèque pilote, un comité des utilisateurs, un service d'assistance technique.

(2) Indexation actuelle par descripteurs

Bibliothèque universitaire, Recherche documentaire, Recherche automatisée, Mode conversationnel, Personnel, Formation professionnelle, France.

(3) Résumé en langue naturelle

Recherche documentaire automatisée en France par les bibliothèques universitaires. Accès en mode conversationnel aux bases de données. Formation professionnelle du personnel affecté à ces activités.

(4) Indexation avec V1: écriture par le documentaliste

1. Recherche t1 documentaire a1 automatisée a4 par les bibliothèques a1 b universitaires a1 en France a1.

2. Accès t1 de ces bibliothèques a1 en mode conversationnel a1 aux bases de données t0.

3. Formation t1 professionnelle a1 du personnel de ces bibliothèques t3 affecté à cette recherche a1.

(5) Sortie en langue naturelle produite par la machine

1. Recherche documentaire automatisée par les bibliothèques universitaires en France.

2. Accès de ces bibliothèques en mode conversationnel aux bases de données.

3. Formation professionnelle du personnel de ces bibliothèques affecté à cette recherche.

EXEMPLE 5

- (1) B.S. 101, 1978, n°292.

78-101-292. SAUARD (R.). (Bibl. natl. Québec, Montreal, Qué.). La microédition à la bibliothèque nationale du Québec. *Micro Notes*, Canada, (1977), 5, no 2, 8-9. bibl. (5 réf.).
Description du service de microphotographie de la bibliothèque, servant à la conservation et à la microédition. Le support actuel est le microfilm, mais la microfiche est en prévision.

- (2) Indexation actuelle par descripteurs

Bibliothèque nationale, Microreproduction, Microfilm, Conservation document, Microédition, Canada.

- (3) Résumé en langue naturelle

Service de microreproduction par microfilm d'une bibliothèque nationale au Canada, destiné à la conservation et à la microédition.

- (4) Indexation avec V1: écriture par le documentaliste

1.1 Service de microreproduction t3 par microfilm a1 d'une bibliothèque a1 b nationale a1 au Canada a1. 2 Conservation t1 et microédition t1. 3 Ce service t3 avec l'objectif de r cette conservation t1 et microédition t1.

- (5) Sortie en langue naturelle produite par la machine

1. Service de microreproduction par microfilm d'une bibliothèque nationale au Canada. Conservation et microédition. Ce service avec l'objectif de cette conservation et de cette microédition.

EXEMPLE 6

- (1) B.S. 101, 1978, n°366.

78-101-366. FARRADANE (J.), GULUTZAN (P.). (Univ. Western Ontario, Can.). A test of relational indexing integrity by conversion to a permuted alphabetical index.

Internation. Classif., Germ., (1977), 4, no 1, 20-25, bibl. (4 réf.).
L'indexation par relation a été testée pour la déformation du sens lors de l'indexation. On rappelle les principes de ce type d'indexation qui consiste à utiliser un certain type d'opérateurs pour représenter des relations d'association ou de discrimination entre les termes, un diagramme est ainsi formé par des chaînes de mots avec des opérateurs intercalés. Un programme a été élaboré pour convertir ces diagrammes en un index alphabétique permuté. Les résultats ont confirmé que ce type d'indexation ne déforme pas le sens d'un contenu de manière essentielle.

- (2) Indexation actuelle par descripteurs

Indexation, Indicateur rôle, Evaluation, Traitement automatisé, Index permuté.

- (3) Résumé en langue naturelle

Evaluation de l'indexation par indicateur de rôle au moyen de la conversion automatique de cette indexation en un index permuté.

- (4) Indexation avec V1: écriture par le documentaliste

1. Evaluation t1 de l'indexation t1 par indicateur de rôle a1.
- 2.1 Conversion t1 automatique a1 de cette indexation t1.
- 2 Confection t1 automatique a1 par cette conversion a1 d'un index t0
- 1 2 3 4
- permuté a4 (2,4,5,3).
- 5
3. Cette évaluation t1 au moyen de r cette confection t1.

- (5) Sortie en langue naturelle produite par la machine

1. Evaluation de l'indexation par indicateur de rôle.
2. Conversion automatique de cette indexation. Confection automatique d'un index permuté par cette conversion.
3. Cette évaluation au moyen de cette confection.

EXAMPLE 7

- L'étude a pour objectif d'analyser l'impact de l'informatisation sur la cellule de documentation et sur les utilisateurs. On présente la méthodologie de l'étude, la description des tâches du service, et l'étude des comportements du personnel, l'enquête menée parmi les utilisateurs.

Centre documentation, Activité documentaire, Personnel, Traitement automatisé, Utilisateur, Enquête, Urbanisme, France.

Conséquences de l'automatisation sur un centre de documentation sur l'urbanisme en France et sur ses utilisateurs. Activités documentaires de ce centre. Comportement de son personnel. Enquête menée auprès des utilisateurs de ce centre.

1. Centre de documentation t3 en France a1 sur l'urbanisme a1.
- 2.1 Automatisation t1 de ce centre t3. 2 Utilisateurs t3 de ce centre t3. 3 Cette automatisation t1 influençant r ce centre t3.
4. & 3= ces utilisateurs t3.
3. Activités t0 documentaires a1 de ce centre a3.
4. Comportement t0 du personnel de ce centre a3.
5. Enquête t1 auprès de ces utilisateurs t3.

1. Centre de documentation en France sur l'urbanisme.
2. Automatisation de ce centre. Utilisateurs de ce centre. Influence de cette automatisation sur ce centre. Influence de cette automatisation sur ces utilisateurs.
3. Activités documentaires de ce centre.
4. Comportement du personnel de ce centre.
5. Enquête auprès de ces utilisateurs.

EXEMPLE 8

- (1) B.S. 101, 1978, n°379.

78-101-379. SCHULZ (M. J.). Une banque de données terminologiques au service du traducteur possibilités d'interrogation dans le système TEAM.

In: *Franchir barrière linguist. Congr. eur. syst. réseaux doc. 3; Luxembourg; 1977*, München, Verlag dok., (1977), 1, 153-191, bibl. (1 p. 1/2).

Le système TEAM, constitué par une banque de données terminologiques multilingues, a été développé par le service de traduction de la société Siemens A.G. pour mettre à la disposition du traducteur des aides automatiques. L'étude examine les données terminologiques stockées, ainsi que quelques problèmes posés par leur récupération et étudie les possibilités d'interrogation: par dialogue et terminaux, ou traitement automatique de textes lisibles par ordinateur pour établir des glossaires.

- (2) Indexation actuelle par descripteurs

Vocabulaire, Banque de données, Traduction, Multilinguisme, Recherche automatisée, Mode conversationnel.

- (3) Résumé en langue naturelle

Banque de données terminologiques multilingues pour assister automatiquement le traducteur. Données terminologiques stockées et récupération de ces données. Interrogation de cette banque, en mode conversationnel ou par traitement automatique des textes en vue de la confection automatique de glossaires.

- (4) Indexation avec V1: écriture par le documentaliste

1. Banque de données t0 terminologiques a1 multilingues a1.
 2.1 Assistance t1 automatique a1 au traducteur t3. 2 Cette banque t0 avec l'objectif de r cette assistance t1.
 3. Données t0 terminologiques a1 stockées dans cette banque a1.
 4. Recherche t1 de ces données t0.
 5. Interrogation t1 en mode conversationnel a1 de cette banque t0.
 6.1 Traitement t1 automatique a1 des textes t0. 2 Confection t1 automatique a1 de glossaires t0. 3 Ce traitement t1 avec l'objectif de r cette confection t1. 4 Interrogation t1 automatique a1 par ce traitement a1 de cette banque t0 (2,4,3).
 1 2
 3 4

- (5) Sortie en langue naturelle produite par la machine

1. Banque de données terminologiques multilingues.
 2. Assistance automatique au traducteur. Cette banque avec l'objectif de cette assistance.

3. Données terminologiques stockées dans cette banque.
4. Recherche de ces données.
5. Interrogation en mode conversationnel de cette banque.
6. Traitement automatique des textes. Confection automatique de glossaires. Ce traitement avec l'objectif de cette confection. Interrogation automatique de cette banque par ce traitement.

EXEMPLE 9

(1) B.S. 101, 1978, n°139.

78-101-139. DILGER (M.). (Aachen). Dokumentation über die Beziehungen der nordrhein-westfälischen Universitätsbibliotheken zu den bibliothekarischen Einrichtungen Belgiens und der Niederlande. (Documentation sur les relations des bibliothèques universitaires de Rhénanie du Nord-Westphalie avec les bibliothèques de Belgique et des Pays-Bas).

Verb. Bibl. Land. Nordrhein-Westfalen, Mittbl., Dtsch., (1977), 27, no 2, 155-162.

Ces relations consistent surtout en opérations d'échanges de documents et de prêt. Sous la forme de tableaux on signale, pour chacune des bibliothèques universitaires du Land, les bibliothèques de Belgique et des Pays-Bas avec lesquelles des relations: 1) d'échanges de documents, 2) de prêt, sont entretenues. Les autres formes de relations et de coopération entre les bibliothèques sont aussi signalées là où elles existent.

(2) Indexation actuelle par descripteurs

Bibliothèque universitaire, Allemagne (république fédérale), Coopération, International, Bibliothèque, Belgique, Pays-Bas, Echange document, Prêt.

(3) Résumé en langue naturelle

Coopération des bibliothèques universitaires d'Allemagne (république fédérale) avec les bibliothèques de Belgique et des Pays-Bas, notamment par échange de documents et prêt interbibliothèque.

(4) Indexation avec V1: écriture par le documentaliste

1.1 Bibliothèques t0 universitaires a1 d'Allemagne (république fédérale) a1. 2 Bibliothèques t0 de Belgique a1. 3 & 2= des Pays-Bas a1.
 1 2
 4 Ces(1.1) bibliothèques t0 en coopération avec r ces(1.2) bibliothèques t0. 5 & 3= ces(1.3) bibliothèques t0.

2.1 Echange t1 de documents t0. 2 Cette coopération t1 internationale a1 par cet échange a1.
 1 2
 3

3. & 3= par prêt a1 b interbibliothèque a1.

(5) Sortie en langue naturelle produite par la machine

1. Bibliothèques universitaires d'Allemagne (république fédérale). Bibliothèques de Belgique. Bibliothèques des Pays-Bas. Ces bibliothèques universitaires d'Allemagne (république fédérale) en coopération avec ces bibliothèques de Belgique. Ces bibliothèques universitaires d'Allemagne (république fédérale) en coopération avec ces bibliothèques des Pays-Bas.

2. Echange de documents. Cette coopération internationale par cet échange.

3. Cette coopération internationale par prêt interbibliothèque.

EXEMPLE 10

(1) B.S. 101, 1978, n°2950.

78-101-2950. HOELL (C.). (Berlin). *Ergebnisse und Perspektiven der Bibliotheksarbeit im Zuständigkeitsbereich des Ministeriums für Volksbildung*. (Résultats et perspectives du travail des bibliothèques dans le secteur du ressort du Ministère de l'éducation populaire).

Zbl. Bibl.-Wes., Dtsch., (1977), 91, no 12, 559-564, rés. russe angl. fr., bibl. (4 réf.).

On rapporte les expériences relatives au rôle et à l'activité des bibliothèques spécialisées en pédagogie, qui dépendent d'établissements d'enseignement supérieur ou autres organismes spécialisés dans la formation de futurs enseignants ou éducateurs, depuis l'entrée en vigueur de l'ordonnance de 1968 sur les bibliothèques. On met en lumière la coopération qui existe entre ces bibliothèques et les organismes d'information et de documentation en pédagogie, et également les bibliothèques publiques. On note en particulier l'étroite coopération entre la bibliothèque centrale de pédagogie et l'Institut central pour le perfectionnement des enseignants et éducateurs de Ludwigsfelde; on souligne aussi la coopération avec les bibliothèques et les organismes spécialisés de l'URSS et des pays socialistes. Enfin on montre de quelle manière est assuré le perfectionnement de la qualification professionnelle et politique des personnels des bibliothèques.

(2) Indexation actuelle par descripteurs

Bibliothèque enseignement supérieur, Bibliothèque spécialisée, Education, Organisme documentation, Coopération, Formation permanente, Personnel, National, International, Allemagne (R.D.A.).

(3) Résumé en langue naturelle

Activités des bibliothèques d'Allemagne (R.D.A.), spécialisées en éducation, qui dépendent d'établissements d'enseignement supérieur ou autres organismes spécialisés dans la formation des enseignants. Coopération de ces bibliothèques avec les organismes d'information et de documentation en pédagogie et également avec les bibliothèques publiques. Coopération de ces bibliothèques allemandes avec les bibliothèques et les organismes spécialisés de l'U.R.S.S. et des pays socialistes. Formation professionnelle et politique des personnels de ces bibliothèques spécialisées allemandes.

(4) Indexation avec V1: écriture par le documentaliste

1.1 Bibliothèques t0 spécialisées a4 b en éducation al. 2 Ces bibliothèques t0 en Allemagne (R.D.A.) al dépendant des organismes de
 1 2
 l'enseignement a3 b supérieur al. 3 & 3-4= dépendant d'autres orga-
 3 4
 nismes al b spécialisés a4 b en éducation al.

2.1 Organismes d'information t3 et de documentation t3 en éducation a1. 2 Bibliothèques t0 publiques a1. 3 Ces(1.1) bibliothèques t0¹ en coopération avec r ces(2.1) organismes t3. 4 & 3= ces(2.2) bibliothèques²₃.

3.1 Bibliothèques t0 et organismes t0 de l'U.R.S.S. a1 spécialisés a4 b en éducation a1. 2 Ces(1.1) bibliothèques [spécialisées allemandes] t0 en coopération avec r ces(3.1) bibliothèques [et organismes de l'U.R.S.S.] t0.

4. Coopération t1 internationale a1 de ces(1.1) bibliothèques [spécialisées allemandes] a1.

5. Formation t1 professionnelle a1 et politique a1 des personnels de ces(1.1) bibliothèques [spécialisées allemandes] t3.

(5) Sortie en langue naturelle produite par la machine

1. Bibliothèques spécialisées en éducation. Ces bibliothèques en Allemagne (R.D.A.) dépendant des organismes de l'enseignement supérieur. Ces bibliothèques en Allemagne (R.D.A.) dépendant d'autres organismes spécialisés en éducation.

2. Organismes d'information et de documentation en éducation. Bibliothèques publiques. Ces bibliothèques spécialisées en éducation en coopération avec ces organismes d'information et de documentation en éducation. Ces bibliothèques spécialisées en éducation en coopération avec ces bibliothèques publiques.

3. Bibliothèques et organismes de l'U.R.S.S. spécialisés en éducation. Ces bibliothèques spécialisées allemandes en coopération avec ces bibliothèques et organismes de l'U.R.S.S.

4. Coopération internationale de ces bibliothèques spécialisées allemandes.

5. Formation professionnelle et politique des personnels de ces bibliothèques spécialisées allemandes.